

(12) NACH DEM VERTRAG ÜBER DIE INTERNATIONALE ZUSAMMENARBEIT AUF DEM GEBIET DES
PATENTWESENS (PCT) VERÖFFENTLICHTE INTERNATIONALE ANMELDUNG

(19) Weltorganisation für geistiges Eigentum
Internationales Büro



(43) Internationales Veröffentlichungsdatum
13. Juni 2002 (13.06.2002)

PCT

(10) Internationale Veröffentlichungsnummer
WO 02/47026 A2

(51) Internationale Patentklassifikation⁷: G06N 3/00

49, 83624 Otterfing (DE). HARBECK, Nadia [DE/DE];
Palnkamer Strasse 49, 83624 Otterfing (DE).

(21) Internationales Aktenzeichen: PCT/EP01/14411

(22) Internationales Anmeldedatum:
7. Dezember 2001 (07.12.2001)

(74) Anwalt: WEIGELT, Udo; Grünecker, Kinkeldey,
Stockmaier & Schwanhäusser, Anwaltssozietät, Maximil-
ianstrasse 58, 80538 München (DE).

(25) Einreichungssprache: Deutsch

(26) Veröffentlichungssprache: Deutsch

(30) Angaben zur Priorität:
100 60 928.7 7. Dezember 2000 (07.12.2000) DE

(71) Anmelder und

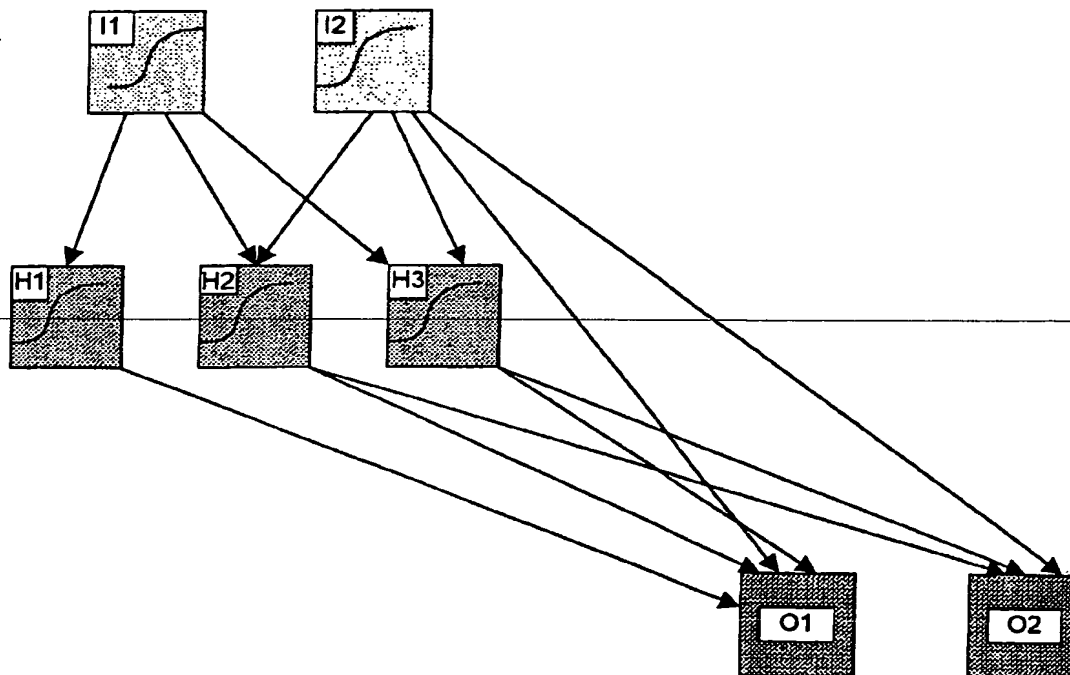
(72) Erfinder: KATES, Ronald, E. [DE/DE]; Palnkamer Str.

(81) Bestimmungsstaaten (national): AE, AG, AL, AM, AT,
AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR,
CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE,
GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR,
KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK,
MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU,
SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US,
UZ, VN, YU, ZA, ZM, ZW.

[Fortsetzung auf der nächsten Seite]

(54) Title: METHOD FOR DETERMINING COMPETING RISKS

(54) Bezeichnung: VERFAHREN ZUR ERMITTLUNG KONKURRIERENDER RISIKEN



(57) Abstract: The invention relates to a method for determining competing risks for objects after an initial event based on already measured or otherwise objectifiable training data sets, in which a number of signals obtained from an adaptive system are combined in an objective function in such a manner that the adaptive system can identify or predict the underlying probabilities of the respective competing risks.

[Fortsetzung auf der nächsten Seite]

WO 02/47026 A2

BEST AVAILABLE COPY



(84) **Bestimmungsstaaten** (*regional*): ARIPO-Patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), eurasisches Patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), europäisches Patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI-Patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Veröffentlicht:

— *ohne internationalen Recherchenbericht und erneut zu veröffentlichen nach Erhalt des Berichts*

Zur Erklärung der Zweibuchstaben-Codes und der anderen Abkürzungen wird auf die Erklärungen ("Guidance Notes on Codes and Abbreviations") am Anfang jeder regulären Ausgabe der PCT-Gazette verwiesen.

(57) **Zusammenfassung:** Die Erfindung betrifft ein Verfahren zur Ermittlung konkurrierender Risiken für Objekte nach einem Anfangsereignis auf Grundlage bereits gemessener oder sonst objektivierbarer Trainingsdatensätze, in welchem mehrere aus einem lernfähigen System gewonnene Signale in einer objektiven Funktion so kombiniert werden, dass das lernfähige System die zugrunde liegenden Wahrscheinlichkeiten der jeweiligen konkurrierenden Risiken erkennen oder vorhersagen kann.

Verfahren zur Ermittlung konkurrierender Risiken

Gebiet der Erfindung

Die Erfindung betrifft ein Verfahren zur Ermittlung konkurrierender Risiken nach einem Anfangsereignis mit Hilfe lernfähiger Systeme auf Grundlage bereits gemessener oder sonst objektivierbarer Daten (Trainingsdaten).

Stand der Technik

Lernfähige Systeme wie neuronale Netze werden zunehmend zur Risikoabschätzung eingesetzt, weil sie in der Lage sind, komplexe, nicht von vornherein bekannte Zusammenhänge zwischen erhobenen Faktoren und Ausgängen zu erkennen und darzustellen. Durch diese Fähigkeit können sie zuverlässigere bzw. genauere Abschätzungen von Risikowahrscheinlichkeiten liefern als herkömmliche Verfahren, welche von einer speziellen Gestalt des Zusammenhangs wie etwa einer linearen Abhängigkeit, ausgehen müssen.

Im Bereich medizinischer Anwendungen, beispielsweise bei der Behandlung einer Krebserkrankung, ist es bekannt, lernfähige Systeme wie neuronale Netze oder rekursive Partitionierung (wie die bekannte Methode CART, „Classification and Regression Trees“, siehe dazu z. B. L. Breiman *et al.*, „Classification and Regression Trees“, Chapman and Hall, New York (1984)) zur Bestimmung der Risikowahrscheinlichkeit eines Ereignisses auch bei zensierten Daten einzusetzen. (Von einem zensierten Datensatz spricht man, wenn bis zum letzten Beobachtungszeitpunkt das Ereignis noch nicht unbedingt eingetroffen ist.) Als Beispiel für die Anwendung lernfähiger Systeme bei einer Krebserkrankung dient die Bestimmung der Risikowahrscheinlichkeit (etwa für eine erneute Erkrankung (Rezidiv)) im Anschluß an der Primärbehandlung zur Unterstützung der Therapieentscheidung.

Die Faktoren der Datensätze umfassen eine Reihe von objektivierbaren Kenngrößen, auf deren Werte eine das lernfähige System bedienende Person keinen Einfluß hat. Im Falle eines primären Mammakarzinoms beinhalten diese Kenngrößen zum Beispiel

- Alter zum Zeitpunkt der Operation,
- Anzahl der befallenen Lymphknoten,
- Laborwert des Faktors uPA,
- Laborwert des Faktors PAI-1,
- Kennwert für die Tumorgroße,
- Laborwert des Estrogenrezeptors,
- Laborwert des Progesteronrezeptors.

Die Art der tatsächlich verwendeten Therapie kann als Angabe erfaßt werden, so dass auch der Zusammenhang zwischen Therapie und Ausgang erkannt wird.

Die Werte werden auf einem geeigneten Speichermedium zwischengespeichert und dem lernfähigen System zugeführt. Die einzelnen Angaben sind in der Regel allerdings zum einen mit einer Unschärfe, analog einem Signalrauschen, behaftet. Aus diesen verrauschten Einzelsignalen ist es Aufgabe des lernfähigen Systems, veredelte Signale zu bilden, die im Rahmen einer geeigneten Wahrscheinlichkeitsdarstellung zu einer Risikoabschätzung führen können.

Die Lernfähigkeit eines neuronalen Netzes auch für nichtlineare Zusammenhänge ist eine Konsequenz der Architektur und der Funktionsweise. Ein sog. „Multilayer Perzeptron“ (in der Fachliteratur immer als „MLP“ abgekürzt) enthält etwa eine Eingangsschicht, eine versteckte Schicht, und eine Ausgangsschicht. Die im neuronalen Netz vorhandenen „versteckten Knoten“ erfüllen die Aufgabe, ein Signal für die Wahrscheinlichkeit komplexer interner Prozesse zu erzeugen. Sie können somit über die zugrunde liegenden, aber nicht von vornherein direkt abtastbaren, biologischen Prozesse, welche für den Verlauf einer Erkrankung letztendlich ausschlaggebend sein werden, Aufschluß geben.

Interne biologische Prozesse können parallel zu einander mit unterschiedlichen Raten stattfinden und auch miteinander in Wechselwirkung treten. Lernfähige Systeme können auch solche internen, nicht direkt beobachtbaren Prozesse erkennen und darstellen, wobei sich die Güte dieser Erkennung nachträglich und indirekt durch die Güte der Vorhersage der tatsächlich beobachteten Ereignisse bemerkbar macht.

Durch rekursive Partitionierung (etwa CART) entstehen *Zuordnungen*, die in ihrer Fähigkeit zur Darstellung komplexer interner Zusammenhänge analog zu den Fähigkeiten der neuronalen Netze sind.

Der Verlauf einer Erkrankung kann zu unterschiedlichen kritischen Ereignissen führen, deren Vorbeugung eventuell unterschiedliche Therapieansätze erfordert. Im Falle des ersten Rezidivs bei Brustkrebs ist eine eindeutige Klassifikation der Befunde etwa in die gegenseitig exklusiven Ausprägungen

1. „Femmetastasen in Knochengewebe“,
2. „Femmetastasen, jedoch kein Befund im Knochengewebe“
3. „Loko-regionales“ Rezidiv

möglich.

Da aber der weitere Verlauf der Erkrankung nach einem dieser Befunde auch im Hinblick auf die Wahrscheinlichkeiten der übrigen Ausprägungen der Rezidivs beeinflusst werden kann, ist es im Rahmen der statistischen Behandlung solcher Daten oft sinnvoll, nur das Erstrezidiv zu untersuchen. Beispielsweise für eine Brustkrebspatientin, die 24 Monate nach der Primäroperation ein Lokalrezidiv erleidet und nach 48 Monaten den Befund „Knochenmetastase“ hat, ist im Hinblick auf „Erstrezidiv“ nur Kategorie 3 relevant. Die Nachbeobachtung Knochenmetastase wird im diesem Rahmen nicht verwendet, d.h., die Patientin ist im Hinblick auf den Befund 1 als „zensiert“ zu betrachten, sobald ein anderer Befund (hier Lokalrezidiv) festgestellt worden ist.

Konkurrierende Risiken können auch dadurch entstehen, dass ein Patient etwa an einer völlig anderen Erkrankung oder an einer Nebenwirkung der Behandlung verstirbt, so dass das Risiko der den Arzt interessierenden Ausprägung verborgen bleibt.

Es ist für Experten zwar relativ offensichtlich, dass eine *ausschließliche Klassifikation mit Zensierungsvorschrift* die Trainingsdaten in eine Gestalt so abbilden kann, dass für jeden möglichen Ausgang je ein neuronales Netz oder je ein Klassifikationsbaum durch rekursive Partitionierung nach Stand der Technik trainiert werden kann. Im Beispiel mit den Ausgängen

1 bis 3 müßte man hierfür drei vollkommen unabhängige neuronale Netze bzw. drei verschiedene Entscheidungsbäume trainieren.

Ein Problem bei dieser Nutzung des Standes der Technik besteht darin, dass die Erkennung einer möglichen Aussagekraft interner Knoten hinsichtlich *eines* der Erkrankungsausgänge für die Erkennung dessen Aussagekraft hinsichtlich der *übrigen* Erkrankungsausgänge verlorenggeht. In Wirklichkeit aber könnte ein interner biologischer Prozess, der durch interne Knoten in einem neuronalen Netz erkannt worden ist, Beiträge zu mehreren beobachtbaren Ausgängen liefern, wenn auch mit unterschiedlicher Gewichtung. Beispielsweise ist die biologische „Invasionsfähigkeit“ eines Tumors von unterschiedlicher, jedoch signifikanter Bedeutung für Fernmetastasen bzw. für lokale Rezidive. Die unabhängig trainierten Netze müssen die Aussagekraft eines durch den Knoten dargestellten internen Zusammenhang eigenständig „entdecken“.

Es versteht sich, dass die Anzahl der tatsächlichen Ereignisse, die einem lernfähigen System zur Verfügung stehen, analog zur Mächtigkeit eines statistischen Systems, die Erkennungsgüte mitbestimmen. In medizinischen Anwendungen ist diese Anzahl meist begrenzt. Infolgedessen ist die Wahrscheinlichkeit, dass sich interne Prozesse knapp hinsichtlich eines der Ausgänge bemerkbar machen, jedoch nicht hinsichtlich der übrigen Faktoren, relativ groß. In diesem Fall geht die potentielle Aussagekraft zur Unterscheidung der Faktoren, sowie das biologische Erklärungspotential der internen Knoten auch im Hinblick auf die weiteren Ausgänge, verloren.

Da Therapien auch Nebenwirkungen haben, ist etwa eine Verringerung des Risikos einer Ausprägung einer Erkrankung auf Kosten einer Erhöhung eines anderen Risikos typisch für den medizinischen Entscheidungskontext. Hierfür ist die Notwendigkeit beim gegenwärtigen Stand der Technik, hinsichtlich jedes einzelnen Risikos ein völlig neues neuronale Netz trainieren zu müssen, unbefriedigend.

Faktoren, deren Wirkung auf die Wahrscheinlichkeit der Ausgänge zeitlich variabel sind, können nach Stand der Technik durch verschiedene Knoten in der Ausgangsschicht dargestellt werden, denen verschiedene Zeitabhängigkeiten (etwa durch die bekannte Technik der „funktionalen Polynomen“) zugeordnet werden. Eine zeitlich variable Aussage zur Ereignisdichte ist beim Stand der Technik zwar möglich, jedoch ist das Problem der

konkurrierenden Risiken nicht so formulierbar, dass die Bestimmung einer zeitlich variablen Aussage nicht beeinträchtigt wird.

Angesichts der Nachteile des Standes der Technik liegt der Erfindung die Aufgabe zugrunde, ein Verfahren bereitzustellen, mit dem konkurrierenden Risiken in ihrem logischen bzw. ursächlichen Zusammenhang erfaßt, identifiziert und dargestellt werden können, insbesondere so, dass die Bestimmung einer zeitlich variablen Aussage nicht beeinträchtigt wird.

Beschreibung der Erfindung

Diese Aufgabe wird durch das Verfahren gemäß Patentanspruch 1 gelöst.

Durch das erfindungsgemäße Verfahren können den konkurrierenden Risiken durch das lernfähige System geeignete Kennwerte zugeordnet werden. Diese Kennwerte sollen die Berechnung der bedingten Wahrscheinlichkeit pro Zeiteinheit für das Eintreten des jeweiligen Ereignisses (unter der Voraussetzung, dass bisher keines der möglichen Endereignisse eingetreten ist) ermöglichen. „Geeignete“ Kennwerte im Sinne der Erfindung können die Eigenschaft haben, dass ein Maximum der statistischen „Likelihood“ betreffend aller Ausgänge angestrebt wird.

Es versteht sich, dass dieses Verfahren in verschiedensten Bereichen, wie z. B. den Ingenieurs- und Wirtschaftswissenschaften, der Biologie oder Medizin, verwendet werden kann. Im Bereich der Medizin kann es sich dann bei den Objekten um Patienten handeln, die nach einer ersten Erkrankung, dem Ausgangsereignis, konkurrierenden Risiken unterliegen, eine erneute Erkrankung zu erleiden.

Vorteilhafterweise werden für das Verfahren für die Trainingsdatensätze gemessene oder in sonstiger Weise objektiv erfaßte Daten des Anfangsereignisses und einer Nachbeobachtung bis zu einem vorgegebenen Zeitpunkt verwendet.

Dabei ist es von Vorteil, wenn der letzte Zeitpunkt der Nachbeobachtung explizit in den Trainingsdatensätzen verwendet wird.

Somit kann durch das erfindungsgemäße Verfahren auch die Nutzung anderer Kennwerte im Rahmen eines trainierten lernfähigen Systems ermöglicht werden, solange sich diese Kennwerte analog zur statistischen Likelihood aus den Nachbeobachtungen bilden lassen.

In einer vorteilhaften Ausführung werden bei der Beobachtung einer Versagensausprägungen zu einem Zeitpunkt die anderen Ausprägungen ausgeschlossen. Auf diese Weise kann eine Ausprägung eines Versagens bevorzugt berücksichtigt werden.

Vorteilhafterweise ist die objektive Funktion L in Abhängigkeit einer Funktion P gegeben:

$$L(\mu; \{\mathbf{x}_j, t_j, \delta_{jk}\}) = \prod_{j=1}^n P\left(\left[f_{LS(k, \mathbf{x}_j)}(t_j)\right] \left[S_{LS(k, \mathbf{x}_j)}(t_j)\right], k=1, \dots, K\right),$$

Hier sind mit μ die Parameter des lernfähigen Systems gemeint. („LS“ steht für „lernfähiges System“.) $f_{LS(k, \mathbf{x}_j)}(t_j)$ bezeichnet die Versagensrate der Ausprägung k und $S_{LS(k, \mathbf{x}_j)}(t_j)$ den Erwartungswert des Anteils an Objekten j mit beobachteten Merkmalen \mathbf{x}_j , welche kein Versagen der Ausprägung k zum Zeitpunkt t_j erleiden. P wird aufgrund des logischen Zusammenhangs aus δ_{jk} bestimmt, wobei $\delta_{jk} = 1$, wenn ein Objekt j ein Versagen der Ausprägung k zum Zeitpunkt t_j erlitten hat und sonst $\delta_{jk} = 0$.

Vorteilhafterweise wird

$$L(\mu; \{\mathbf{x}_j, t_j, \delta_{jk}\}) = \prod_{j=1}^n \prod_{k=1}^K \left[f_{LS(k, \mathbf{x}_j)}(t_j)\right]^{\epsilon_{jk}} \left[S_{LS(k, \mathbf{x}_j)}(t_j)\right]^{\psi_{jk}}$$

als objektive Funktion verwendet, wobei ϵ_{jk} und ψ_{jk} aufgrund des logischen Zusammenhangs aus δ_{jk} bestimmt werden.

Es ist von Vorteil, wenn

$$L(\mu; \{\mathbf{x}_j, t_j, \delta_{jk}\}) = \prod_{j=1}^n \prod_{k=1}^K \left[f_{LS(k, \mathbf{x}_j)}(t_j)\right]^{\delta_{jk}} \left[S_{LS(k, \mathbf{x}_j)}(t_j)\right]^{-\delta_{jk}}$$

als objektive Funktion verwendet wird.

In einer bevorzugten Alternative wird ein neuronales Netz als lernfähiges System verwendet. In diesem Fall kann obige objektive Funktion L in Abhängigkeit von P die Form haben

$$L(\mu, \{x_j, t_j, \delta_{jk}\}) = \prod_{j=1}^n P\left(\left[f_{NN(k, x_j)}(t_j)\right] \left[S_{NN(k, x_j)}(t_j)\right] k = 1, \dots, K\right).$$

Von besonderem Vorteil ist es, wenn ein neuronale Netz der Architektur MLP (multilayer Perzeptron) verwendet wird.

In einer anderen bevorzugten Alternative führt das lernfähige System eine rekursive Partitionierung durch, wobei

jedem Objekt einem Knoten zugeordnet wird,

jedem Knoten die Häufigkeiten oder Wahrscheinlichkeiten aller Ausprägungen zugeordnet wird, und

die Partitionierung so durchgeführt wird, dass die objektive Funktion optimiert wird, die diese Häufigkeiten oder Wahrscheinlichkeiten statistisch berücksichtigt.

Vorzugsweise wird das lernfähige System im Rahmen einer Entscheidungshilfe verwendet.

Es ist von Vorteil, wenn den verschiedenen Wahrscheinlichkeitsfunktionen der konkurrierenden Risiken Werte zur Ermittlung einer Strategie zugeordnet werden. Damit kann beispielsweise bei einer medizinischen Anwendung der vorliegenden Erfindung eine Therapiestrategie ermittelt werden.

Im folgenden wird das erfindungsgemäße Verfahren zur Ermittlung konkurrierender Risiken unter Bezugnahme auf die Zeichnung weiter beschrieben. In der Zeichnung zeigen:

Figur 1 eine Darstellung eines neuronalen Netzes in einer Implementierung als MLP,

Figur 2 ein Venn-Diagramm konkurrierender Risiken, und

Figur 3 eine Illustration eines trainierten neuronalen Netzes mit drei konkurrierenden Risiken.

Die nachfolgend beschriebenen Ausführungsformen beziehen sich auf medizinische Anwendungen, was aber nicht als Einschränkung zu verstehen ist.

Die folgende Beschreibung bedient sich der Terminologie der neuronalen Netze der Architektur MLP. Die Anwendung auf andere Architekturen der neuronalen Netze sowie für Regressionsbäume ist aber analog und für Experten unmittelbar ohne weitere Beschreibung nachzuvollziehen.

Insbesondere wird erfindungsgemäß eine zusätzliche Dimension der Ausgangsschicht des lernfähigen Systems eingeführt, wobei

- die zusätzliche Dimension der Ausgangsschicht mindestens zwei Knoten umfaßt
- die Knoten dieser zusätzlichen Dimension den verschiedenen Ausgangsereignissen entsprechen
- jeder Ausgangsknoten einem Signal zugeordnet wird,
- die einzelnen Signale jeweils einer Risikofunktion bezüglich der möglichen Ereignisse zugeordnet werden.
- die Signale der Ausgangsfunktionen zu einem Gesamtsignal kombiniert werden
- das lernfähige System trainiert wird, indem die Werte der Gesamtsignale für alle Datensätze als Objektivfunktion für das System herangezogen werden

Ein auf diese Weise trainiertes System unterstützt den behandelnden Arzt und den Patienten beispielsweise bei der Entscheidung für einen aus mehreren unterschiedlichen Therapieansätzen, indem es feststellt, gegen welche der möglichen Ausprägungen der Rezidivgefahr die Therapie gerichtet werden soll.

Problemdarstellung und Überblick

Das Ziel der individualisierten Patientenprognose mit konkurrierenden Risiken läßt sich mathematisch so auffassen, dass mehrere Funktionen $f_1(\mathbf{x})$ $f_2(\mathbf{x})$ $f_3(\mathbf{x})$, ... mit dem lernfähigen System, hier mit einem neuronalen Netz $NN_1(\mathbf{x})$, $NN_2(\mathbf{x})$, zu approximieren sind. Genauer gesagt, das neuronale Netz schätzt den Erwartungswert $E(y_k|\mathbf{x})$ der stochastischen Variablen y_k bei beobachteten Merkmalen \mathbf{x} :

$$NN_k(\mathbf{x}) \approx f_k(\mathbf{x}) = E(y_k | \mathbf{x}) .$$

Schematisch läßt sich das neuronale Netz zunächst bei der aktuellen Implementierung als MLP im Ausführungsbeispiel wie in Figur 1 darstellen.

Alle Quadrate stellen hier *Neuronen* dar. Die oben im Bild dargestellten Neuronen liefern entweder

- rohe Patientenmerkmale (Beim primären Brustkrebs beispielsweise uPA, PAI-1, Zahl der befallene Lymphknoten, usw.) oder
- aus diesen Merkmalen bereits aufbereitete Größen (z.B. Werte bereinigt um Mittelwert oder Median und normiert durch Standardabweichung der Werteverteilung) oder
- durch Vorkenntnisse oder andere statistische Methoden abgeleitete Größen.

Diese Neuronen bilden zusammen die *Eingangsschicht*.

Die mittleren Neuronen bilden die *interne Schicht*. Es können auch mehrere interne Schichten vorgesehen sein. Jedes interne Neuron verarbeitet die Signale von den Eingangsneuronen und gibt ein Signal weiter. Die mathematische Beziehung zwischen den "Inputs" zu den internen Neuronen und ihren "Outputs" wird durch das Einpendeln von *synaptischen Gewichten* gesteuert.

Die unteren Neuronen liefern Schätzungen für die erwünschten Kenngrößen (z.B. Erwartungswert des Überlebens) und bilden die *Ausgangsschicht*.

Um dem Netz den unterstellten Zusammenhang $f_1(x) f_2(x) f_3(x), \dots$ beizubringen, stehen m Patientinnen zur Verfügung. Jedem Patienten ist ein Datenmuster (x,y) zugeordnet, wobei für konkurrierende Risiken die Ausgangsvariablen y als „Vektoren“ ($y=[y_1, y_2, y_3, \dots]$) zu verstehen sind. Das Netz muß also anhand der Menge der Datenmuster $\{(x^1, y^1), \dots, (x^m, y^m)\}$ die zugrundeliegende Dynamik lernen. Der Index als Superskript bezieht sich auf den Patienten. Beim Lernen findet nun eine Anpassung der synaptischen Gewichte statt.

Die in der Ausführungsform verwendete Architektur besteht in einem klassischen mehrschichtigen Feedforward-Netz. Neuronen sind wie oben beschrieben in Schichten organisiert. Konnektoren bestehen in der Ausführungsform wie folgt

- Eingangsschicht → versteckte Schicht
- Eingangsschicht → Ausgangsschicht
- Versteckte Schicht → Ausgangsschicht

Die Verwendung von Konnektoren Eingangsschicht → Ausgangsschicht ist zweckmäßig aber nicht zwingend zur Funktion der Erfindung, weil sie nicht unbedingt zur Darstellung einer Abbildung $NN(x)$ notwendig sind.

Funktion neuronaler Netze

Neuronen als Funktionen

Jedes Neuron empfängt ein Stimulationssignal S , verarbeitet dieses gemäß einer vorbestimmten Aktivierungsfunktion $F(S)$ und gibt ein entsprechendes Antwortsignal $A = F(S)$ aus, das allen nachfolgenden Neuronen zugeführt wird, welche noch mit ihm verbunden sind. In der Ausführungsform ist die Aktivierungsfunktion der versteckten Schicht der Tangens Hyperbolicus. Die Erfindung kann ebenso bei Verwendung anderer Aktivierungsfunktionen wie etwa der logistischen Funktion eingesetzt werden.

Transformationen und Eingangsneuronen

Die Faktoren werden anfänglich univariat so transformiert, dass sie in einem Intervall der Größenordnung 1 liegen.

$$X_j = \tanh[(x_j - x_{\text{Median}})/x_Q] \quad (1.a)$$

D.h., zuerst wird der Median x_{Median} subtrahiert und eine Skalierung der Werte mit einem Faktor x_Q vorgenommen: Werte über dem Median werden mit dem 75 %-Quantil skaliert, Werte unter dem Median mit dem 25 %-Quantil. Darauf wird noch die Funktion \tanh angewandt.

Die Eingangsneuronen haben eine statische Funktion und werden deshalb als Felder implementiert, welche die so transformierten Werte weiterreichen. Konzeptuell kann man die \tanh -Funktion der Gleichung (1a) als Aktivierungsfunktion der Eingabeschicht ansehen.

Versteckte Neuronen

Gesucht ist der Ausgang des versteckten Knotens h für Patientin j . Zuerst wird überprüft, ob der versteckte Knoten h noch aktiv ist. Falls aktiv, werden die Eingangssignale mit den zugehörigen Gewichten multipliziert um die Summe $w_h \cdot x_j$ zu bilden. Genauer gesagt ist das Signal zum versteckten Knoten h bei Muster j eine gewichtete Summe der Eingänge der Gestalt

$$z_h(j) = \sum_i w_{ih} X_i(j),$$

wobei w_{ih} das Gewicht des Konnektors vom Eingangneuron i zum versteckten Neuron h , und $X_i(j)$ die (skalierte) Antwort des i -ten Eingangsneurons darstellt. Die Antwort des versteckten Neurons h ist

$$Y_h(j) = F_h(z_h(j) - b_h). \quad (2.a)$$

Hierbei ist b_h das Bias des versteckten Neurons h , das rechentechnisch wie jedes andere Gewicht des Netzwerks optimiert wird. Im Ausführungsbeispiel ist die nichtlineare Aktivierungsfunktion F_h der Tangens hyperbolicus.

Ausgangsknoten

Gesucht ist der Ausgang des Ausgangsknotens o für Patientin j . Zuerst wird überprüft, ob der Ausgangsknoten o noch aktiv ist. Konnektoren sowohl von der versteckten Schicht wie auch von der Eingangsschicht sind möglich. Für jeden noch aktiven Konnektor werden die zugehörigen Eingangssignale mit den entsprechenden Gewichten multipliziert.

Das Signal z_o wird zunächst gebildet: Das Bias des Neurons b_o wird abgezogen, und auf dieses Resultat wird die Aktivierungsfunktion des Ausgangsneurons o angewandt. Der Ausgang $O_o(j)$ wird somit zu

$$\begin{aligned} z_o(j) &= \sum_i w_{io}(X_i(j) - c_i) + \sum_h w_{ho} Y_h(j) \\ O_o(j) &= F_o(z_o(j) - b_o) \end{aligned} \quad (2.b)$$

Die Aktivierungsfunktion der Ausgangsschicht wird im Ausführungsbeispiel als die Identitätsfunktion gewählt.

Im Ausführungsbeispiel wird das Gesamtbias im Gegensatz zur versteckten Schicht nicht frei optimiert, sondern wird so gewählt, dass das Mediansignal aller Ausgangsneuronen gleich null ist. Dies ist möglich ohne Einschränkung der Allgemeinheit des Modells. Somit verringert sich die Anzahl der zu optimierenden Parameter um die Anzahl der Biasparameter.

Überlebensanalyse bei konkurrierenden, zeitvariablen Risiken im

Rahmen lernfähiger Modelle

Bezug zum Lernfähigen System

Vorgegeben sei ein Patientenkollektiv mit vorhandenen Kovariaten (prognostischen Faktoren) x_j , die zu einem Anfangszeitpunkt $t=0$ (etwa zum Zeitpunkt der Primäroperation) gemessen wurden, sowie Endpunkten t_j . Bei bekanntem Versagen der Ausprägung k des j -ten Patienten zum Zeitpunkt t_j wird $\delta_{jk} = 1$ ($k=1,2,3,\dots$) definiert, falls zensiert (weiterer Verlauf nach $t=t_j$ unbekannt) wird $\delta_{jk}=0$ definiert.

Sei $S_k(t)$ der Erwartungswert des Anteils der Patienten zu einem Zeitpunkt t , welche kein Versagen der Ausprägung k erleiden, wobei $S_k(\infty) = 0$ und $S_k(0) = 1$. Es ist zweckmäßig, jeweils eine Versagensrate $f_k(t)$ und eine „Hazardfunktion“ $\lambda_k(t)$ gemäß

$$\begin{aligned} f_k(t) &\equiv -\frac{dS_k}{dt} \\ \lambda_k(t) &\equiv -\frac{f_k(t)}{S_k(t)} \end{aligned} \quad (3.a)$$

zu definieren, so dass

$$\lambda_k(t) = -\frac{d}{dt} [\log S_k(t)] \quad (3.b)$$

gilt.

Die Interpretation dieser einzelnen Versagensraten ist wie folgt: Wäre es möglich, die übrigen Ausprägungen ohne Einfluß auf die Ausprägung k zu vermeiden, so würde $f_k(t)$ die beobachtete Versagensrate approximieren. In Wirklichkeit wird $f_k(t)$ zwar nicht direkt beobachtet. Für die Nutzung der Erfindung im Rahmen einer Entscheidungshilfe ist jedoch die Erfassung aller Ausprägungen $f_k(t)$ notwendig, damit die Wirksamkeit einer Verringerung einer Ausprägung im Hinblick auf das Gesamtwohl des Patienten berücksichtigt werden kann.

Bei bekanntem Verlauf der Hazardfunktionen $\lambda_k(t)$ erhält man die $S_k(t)$ durch Integration der Gl. (3.b) mit der Anfangsbedingung $S_k(0) = 1$.

Vom neuronalen Netz erhalten wir zu einem Zeitpunkt t nach der Primäroperation für eine Patientin mit Kovariaten x eine "Hazardfunktion" $\lambda_k(t|x)$, die nun von Kovariaten x abhängt. Als Modell für die Hazardfunktion bei vorgegebenen Kovariaten x setzen wir

$$\lambda_k(t | \mathbf{x}) = \lambda_{k0}(t) h_k(t | \mathbf{x}) \quad (4.)$$

mit

$$h_k(t | \mathbf{x}) = \exp \left[\sum_{l=1}^L B_l(t) NN_{kl}(\mathbf{x}) \right] \quad (5.)$$

Die Funktionen $B_l(t)$ werden dem Problem angepaßt gewählt. Möglich sind hier beispielsweise Spline-Funktionen. In dem Ausführungsbeispiel werden für $B_l(t)$ fraktionale Polynome bevorzugt, d.h. $B_l(t) = t^{(l-1)/2}$.

Somit erhält man

$$\lambda_{0k} \exp \left[\sum_l NN_{kl}(x) B_l(t) \right] = - \frac{d}{dt} \log(S_k(t)) \quad (6.)$$

Hierbei wird in der zweiten Gleichung λ_{0k} als eine Konstante betrachtet. Die Zeitabhängigkeit steckt in den Koeffizienten $B_l(t)$. Dieses Modell ist ein Proportional-hazards-Modell, falls $B_1 = 1$ und alle anderen B_l verschwinden. Abweichungen von „Proportional-hazards“ können durch Berücksichtigung von Termen B_l mit $l > 1$ modelliert werden.

In einer breiten Klasse von Anwendungen wird eine Objektivfunktion der Gestalt

$$L(\mu; \{\mathbf{x}_j, t_j, \delta_{jk}\}) = \prod_{j=1}^n P \left(f_{NN(k, \mathbf{x}_j)}(t_j) \right) \left[S_{NN(k, \mathbf{x}_j)}(t_j) \right] \quad k=1, \dots, K \quad (7.)$$

optimiert, wobei die Schreibweise zum Ausdruck bringt, dass P (zunächst auf eine noch nicht näher spezifizierte Art und Weise) von den jeweiligen Überlebens- bzw. Versagenswahrscheinlichkeiten abhängen darf. Diese Abhängigkeit ist problembedingt und geht aus einem logischen Modell für das Auftreten der verschiedenen Ausprägungen hervor. Eine bevorzugte Klasse von Objektivfunktionen der Gestalt (7.) lassen sich als statistische Likelihoodfunktionen verstehen, wobei für die Ausführungsform

$$L(\mu; \{\mathbf{x}_j, t_j, \delta_{jk}\}) = \prod_{j=1}^n \prod_{k=1}^K \left[f_{NN(k, \mathbf{x}_j)}(t_j) \right]^{\delta_{jk}} \left[S_{NN(k, \mathbf{x}_j)}(t_j) \right]^{1-\delta_{jk}} \quad (8.)$$

gewählt wird.

Die beiden Argumente $f_{NN(k,x)}$ und $S_{NN(k,x)}$ sind eindeutig bestimmt unter der Voraussetzung, dass das neuronale Netz oder das sonstige lernfähige Modell die entsprechenden Werte für die Ausgangsknoten liefert. Dies ist in der Ausführungsform stets der Fall.

Hierbei sind ε_{jk} und ψ_{jk} aufgrund des logischen Zusammenhangs aus δ_{jk} zu bestimmen, wobei $\delta_{jk} = 1$, wenn der Patient j ein Versagen der Ausprägung k zum Zeitpunkt t_j erlitten hat und sonst $\delta_{jk} = 0$. Zensierte Datensätze entsprechen Patienten, die gar kein Versagen erlitten haben, so dass $\delta_{jk} = 0$ für *alle* $k=1,2,3,\dots$. Die funktionale Abhängigkeit vom Modell wird symbolisch durch veränderliche Parameter μ gekennzeichnet. Ein Beispiel zur Bestimmung von ε_{jk} und ψ_{jk} wird in der Folge angegeben.

Die mit μ bezeichneten Parameter sind in der Ausführungsform die Überlebenszeitskalen λ_{0k} und die Gewichte des neuronalen Netzwerkes. Der Index j bezeichnet den Patientendatensatz.

In der Ausführungsform wird das Zeitintegral zur Lösung der Gleichung 6 durch die Standardmethode „Romberg-Integration“ gelöst. Somit können beliebige Zeitabhängigkeiten der Funktionen $B_i(t)$ berücksichtigt werden.

Zu einem Zeitpunkt t sei $S(t)$ der Erwartungswert des Anteils der Patienten, welche *kein Versagen irgendeiner der Ausprägungen* $k=1,\dots,K$ erlitten hat. In der Ausführungsform ist diese Größe durch das Produkt der einzelnen Wahrscheinlichkeiten gegeben:

$$S(t) = \prod_{k=1}^K S_k(t). \quad (9.)$$

Spezifikation der Ausführungsform für ein Beispiel

Für eine vollständige Spezifikation der Ausführungsform müssen nun die Größen ψ_{jk} und ε_{jk} angegeben werden. Im folgenden werden zur Veranschaulichung zwei Fälle der Ausführungsform hinsichtlich dieser Funktionen vollständig spezifiziert, die typisch für die Anwendung der Erfindung im Kontext der konkurrierenden Risiken sind.

Man betrachte eine Erkrankung, bei der das Versagen drei Ausprägungen hat. Der Patient wird im Monat t ($t=1,2, \dots$) beobachtet. Hierbei kann im Monat t irgendeine Kombination der drei Ausprägungen oder gar kein Versagen beobachtet werden, so dass der Patient „zensiert“ ist. Die Situation wird als Venn-Diagramm in der Abbildung 1 veranschaulicht.

Im Falle der Krankheit Mammakarzinom könnten die drei Ausprägungen etwa Knochenmetastasen (B für "bone", $k=1$), sonstige Fernmetastasen (D für „distant“, $k=2$), bzw. Lokal/regional (L für „local“, $k=3$) sein. Ein gleichzeitiges Auftreten aller drei Ausprägungen im Beobachtungsmonat t ist möglich. Es kann aber sein, dass aus klinischen, pharmakologischen, oder datentechnischen Gründen die Nachbeobachtung im Monat t nach folgender Logik angegeben wird:

- *Knochenmetastasen (ja/nein) ?*
 - Falls ja, dann $\varepsilon_{j1}=1 \ \varepsilon_{j2}=0 \ \varepsilon_{j3}=0 \ \psi_{j1}=0 \ \psi_{j2}=0 \ \psi_{j3}=0$
 - Falls nein: *sonstige Fernmetastasen (ja/nein) ?*
 - Falls ja, dann $\varepsilon_{j1}=0 \ \varepsilon_{j2}=1 \ \varepsilon_{j3}=0 \ \psi_{j1}=1 \ \psi_{j2}=0 \ \psi_{j3}=0$
 - Falls nein: *lokal/regional (ja/nein)*
 - Falls ja, dann $\varepsilon_{j1}=0 \ \varepsilon_{j2}=0 \ \varepsilon_{j3}=1 \ \psi_{j1}=1 \ \psi_{j2}=1 \ \psi_{j3}=0$
 - Falls nein, dann $\varepsilon_{j1}=0 \ \varepsilon_{j2}=0 \ \varepsilon_{j3}=0 \ \psi_{j1}=1 \ \psi_{j2}=1 \ \psi_{j3}=1$

In anderen Worten:

Bei dieser Zuordnung wird etwa der Beobachtung „Knochenmetastasen“ eine Priorität eingeräumt, so dass nicht gefragt wird, ob die anderen Ausprägungen zum Zeitpunkt t vorkommen oder nicht. Deshalb ist bei Befund „Knochenmetastasen ja“ der Beitrag zur Likelihoodfunktion (8) der j -ten Patientin gemäß dieser Logik offenbar allein durch den Term $f_{NN(1,j)}$ gegeben (kein Term mit $S_{NN(k,j)}$).

Für den Fall „keine Knochenmetastasen, aber sonstige Fernmetastasen“, geht aus der Zuordnung ein Beitrag $f_{NN(2,j)} \times S_{NN(1,j)}$ hervor.

Für den Fall „weder Knochen- noch sonstige Fernmetastasen, aber lokal/regionale Erkrankung“ geht aus der Zuordnung ein Beitrag $f_{NN(3,j)} \times S_{NN(1,j)} \times S_{NN(2,j)}$ hervor.

Für den Fall „zensiert“ geht aus der Zuordnung ein Beitrag $S_{NN(1,j)} \times S_{NN(2,j)} \times S_{NN(3,j)}$ hervor.

Eine Anwendung der Erfindung für Messungen, bei denen stets das Vorhandensein bzw. die Abwesenheit mehrerer Ausprägungen zum Zeitpunkt t beobachtet und berücksichtigt wird, ist möglich, wenn die obigen Gleichungen mit entsprechenden Gleichungen für die

Wahrscheinlichkeit der gleichzeitigen Beobachtung mehrerer Ausprägungen bei geschätzten Werten der Versagenswahrscheinlichkeiten ersetzt werden.

Aufbau eines neuronalen Netzes zur Ermittlung konkurrierender Risiken

Fig. 1 zeigt den Aufbau eines neuronalen Netzes mit der Architektur MLP. In diesem Fall umfaßt das neuronale Netz

- eine Eingangsschicht mit einer Mehrzahl von Eingangsneuronen N_i (i für „input neuron“)
- wenigstens eine Zwischenschicht mit Zwischenneuronen N_h (h für „hidden neuron“)
- eine Ausgangsschicht mit einer Mehrzahl von Ausgangsneuronen N_o (o für „output neuron“)
- einer Vielzahl von Konnektoren, die jeweils zwei Neuronen unterschiedlicher Schichten miteinander verbinden.

In der Ausführungsform gemäß Fig. 1 wird eine zweidimensionale Ausgangsschicht dargestellt, um die durch die Erfindung vorhandene Möglichkeit zur gleichzeitigen Darstellung zeitlich variabler und auch konkurrierender Risiken zu veranschaulichen. Die vereinfachte Darstellung von nicht zeitlich variablen Risiken bildet hierbei den Sonderfall, bei dem nur die Ausprägungsdimension notwendig ist.

Die Anzahl der anfänglich eingesetzten Eingangs-Neuronen N_i wird üblicherweise entsprechend der Anzahl der für das Patientenkollektiv zur Verfügung stehenden objektivierbaren Angaben gewählt. Nach stand der Technik stehen Verfahren zur Verfügung, welche entweder vorab die Anzahl der Eingangs-Neuronen auf ein für die jeweilige Rechenanlage akzeptierendes Maß automatisch reduzieren oder aber im Laufe der Optimierung überflüssige Eingangs-Neuronen automatisch entfernen, so dass in beiden Fällen die Bestimmung der letztendlich verwendeten Eingangs-Neuronen ohne Eingriff der jeweiligen Bedienungsperson erfolgt.

In der Ausführungsform gemäß Fig. 1 wird die ursprüngliche Anzahl der versteckten Neuronen durch die ursprüngliche Anzahl der Eingangs-Neuronen bestimmt, d.h.,

$$N_h = N_i \quad (10.a)$$

Für diesen Fall stehen nach Stand der Technik Verfahren zur Verfügung, welche eine günstige Vorbelegung der Konnektoren ermöglichen.

Die Neuronen der Ausgangsschicht werden in der Ausführungsform gemäß Fig. 1 sinngemäß in einer zweidimensionalen Matrix mit Indizes

$$J_{\text{time}} = 1, \dots, N_{\text{time}} \quad (10.b)$$

$$J_{\text{key}} = 1, \dots, N_{\text{key}} \quad (10.c)$$

angeordnet, wobei die Anzahl der ursprünglich aktiven Neuronen der Ausgangsschicht gegeben ist durch

$$N_o = N_{\text{time}} \times N_{\text{key}} \quad (10.d)$$

Hierbei bezeichnet der Index J_{key} Signale der jeweiligen Ausprägung, während der Index J_{time} das Signale zur jeweiligen Zeitfunktion (etwa „fraktionelle Polynome“ oder Splinefunktionen) bezeichnet. Ein mit zwei Indizes J_{time} , J_{key} bezeichnetes Ausgangsneuron trägt demgemäß zur Bestimmung des Koeffizients der Zeitfunktion J_{time} beim Risiko für die Ausprägung J_{key} . Die Indizes J_{key} bzw. J_{time} entsprechen in der Ausführungsform sinngemäß den Indizes k bzw. l der Gleichungen 4 bis 7. Hierbei entsprechen N_{key} bzw. N_{time} in der Ausführungsform sinngemäß den Größen K bzw. L dieser Gleichungen.

Für die Anwendung im Rahmen der rekursiven Partitionierung stehen ebenfalls Endknoten zur Verfügung, die üblicherweise in einer eindimensionalen Reihe angeordnet werden. Nach Stand der Technik wird jeder Patient einem solchen Knoten zugeordnet. Ebenfalls nach Stand der Technik wird dem Knoten ein Risiko zugeordnet, das als ein (skalares) Signal betrachtet werden kann. Die Erfindung ordnet nun jedem Endknoten statt eines Skalars ein Vektor mit N_{key} Indizes zu.

Lernen

Für das Ausführungsbeispiel ist es Ziel des Lernens (Trainings), einen möglichst hohen Wert dieser Likelihoodfunktion im Parameterraum zu orten, gleichzeitig aber überflüssige Parameter

nach Möglichkeit zu vermeiden. In der Ausführungsform erfolgt das Lernen durch Initialisierung, Optimierungsschritte und Komplexitätsreduktion wie folgt:

Initialisierung

Univariate Analyse

Bevor das gesamte Netz mit allen Gewichten trainiert wird, ist es vorteilhaft, eine univariate Analyse für jeden Faktor durchzuführen. Diese Analyse hat mehrere Anwendungen:

- Die univariate Stärke der Faktoren bzw. deren einzelne prognostische Güte steht für den Vergleich mit dem vollständigen Netz zur Verfügung.
- Univariate Analyse dient zur Bestimmung einer Rangfolge der Faktoren für den Fall, dass weniger Eingangsknoten als Faktoren vorhanden sind.
- Eine Vorbesetzung der Gewichte, die nichtlineare Konfigurationen begünstigt, oder wenigstens nicht benachteiligt, läßt sich aus den univariaten Analysen bilden (siehe unten).

Zunächst wird ein exponentielles Überlebensmodell mit dem einzigen Parameter λ_0 bestimmt. Dieses Modell dient zur Initialisierung und in der nachfolgenden Analyse außerdem zur Kontrolle.

Lineare univariate Modelle

Der nach Gl. (1a) transformierte Wert des j-ten Faktors X_j wird als einzelner Eingang in ein „Netzwerk“ betrachtet, das aus genau einem linearen Konnektor von diesem Eingangsneuron zu einem Ausgangsknoten besteht (d.h., ohne versteckte Knoten). Die Zeitabhängigkeit dieses Ausgangsknotens entspricht dem „Proportional-Hazards-Modell“ ($K=1$) für zensierte Daten. Das daraus resultierende Modell hat nur noch zwei freie Parameter: der Zeitparameter (λ_0) und das Gewicht des Konnektors. Diese werden optimiert und zusammen mit der Güte (Likelihood) und der Signifikanz für nachfolgende Zwecke in einer Tabelle gespeichert.

Nichtlineare univariate Modelle

Als nächstes wird jeweils für jeden Faktor eine Anpassung mit vier Parametern an ein nichtlineares univariates Modell vorgenommen. Hier wird die Transformation des j-ten Faktors mit dem Ergebnis X_j als „Eingangsneuron“ betrachtet. Das univariate „Netzwerk“ besteht nun aus diesem einen Eingangsneuron, einem einzigen versteckten Neuron, und einem einzigen

Ausgangsneuron (ohne linearen Konnektor zwischen Eingangs- und Ausgangsneuron). Die Zeitabhängigkeit dieses Ausgangsknotens entspricht wie oben dem „Proportional-Hazards-Modell“ ($K=1$) für zensierte Daten.

Die vier Parameter entsprechen jeweils der Zeitkonstante (λ_0), dem Gewicht und dem Bias zur versteckten Schicht, sowie dem Gewicht zur Ausgangsschicht. Diese werden optimiert und zusammen mit der Güte (Likelihood) und der Signifikanz für nachfolgende Zwecke in einer Tabelle gespeichert.

Rangfolge der Eingangsvariablen

Nachdem für jeden Faktor die univariaten Modelle bestimmt worden sind, wird die Rangfolge der univariat signifikanten Faktoren gemäß der Beträge der linearen Gewichte bestimmt. Die Numerierung der Eingangsknoten für die nachfolgende Analyse entspricht dieser Rangfolge. Für den Fall, dass weniger Eingangsknoten als Faktoren zur Verfügung stehen, erlaubt diese Prozedur eine objektive Vorselektion der "wichtigsten" Faktoren.

Vorbesetzung der Gewichte

Für die Optimierung des Netzes (Trainieren) müssen zunächst Anfangswerte für die Gewichte gesetzt werden. Hierbei wird eine Vorbesetzung mit dem genauen Wert Null nicht angestrebt. In der Ausführungsform werden zunächst die Gewichte der linearen Konnektoren wie üblich mit kleinen Werten vorbesetzt. Der Zeitparameter wird mit dem aus dem 1-parametrischen Modell bestimmten Wert λ_0 vorbesetzt. Die Anzahl der versteckten Knoten H wird gleich der Anzahl der Eingangsknoten J gewählt. Nun wird der Konnektor von Eingangsneuron j zum versteckten Neuron mit dem gleichen Index $h=j$ zunächst mit dem unter "nichtlineare univariate Modelle" bestimmten Gewicht vorbesetzt. Der entsprechende Bias wird analog mit dem so bestimmten Bias vorbesetzt. Diese beiden Größen werden dann um einen kleinen zufälligen Betrag versetzt. Daher entspricht jeder Ausgang eines versteckten Knotens etwa dem univariat optimalen nichtlinearen Wert.

Für jeden versteckten Knoten h steht der Wert des aus der univariaten Optimierung gewonnenen Gewichts, die wir hier als w_{h1} bezeichnen, zum ersten Neuron der Ausgangsschicht ebenso zur Verfügung. Um nun die Gewichte zur Ausgangsschicht zu initialisieren, werden die Größen w_{h1} , $h=1, \dots, H$ mit H Zufallszahlen gewichtet. Man wählt in der Ausführungsform dazu H Zahlen aus einer Gleichverteilung auf $[0,1]$ und dividiert jede Zahl durch die Summe. Anschließend werden diese und alle übrigen Konnektoren (d.h., Gewichte

von der versteckten Schicht zu Neuronen der Ausgangsschicht mit $k=2$, usw.) um einen kleinen zufälligen Betrag versetzt.

Eine zweite, für neuronale Netze üblichere Möglichkeit der Initialisierung besteht darin, allen Konnektoren kleine, zufällige Gewichte zuzuordnen. Damit befinden sich am Anfang der Optimierung alle Verknüpfungen, auch die über die versteckte Schicht, im linearen Bereich. Für kleine Argumente ist die "Aktivierungsfunktion" fast linear, z.B., $\tanh(x) \approx x$ für kleines x .

Lineare Statistik der Eingangsfaktoren

In der Ausführungsform wird die Kovarianzmatrix aller Eingangsfaktoren berechnet und gespeichert. Bestimmt wird auch eine lineare Regression jedes Faktors auf alle andere Faktoren: $X_2 \approx A X_1 + B$. Eigenvektoren und Eigenwerte der Kovarianzmatrix werden berechnet und protokolliert. Die linearen Beziehungen werden in der Ausführungsform für die verschiedenen Ausdünnungsverfahren verwendet.

Zuordnung der Patientendaten in Trainings- und Validierungsmengen

Für ein lernfähiges System ist es üblich, die vorhandenen Muster zufällig in Trainings-Validierungs- und Generalisierungsmengen zu unterteilen. Beispielsweise kann der Benutzer Prozentsätze (auch null) aller Muster vorgeben, die für die Validierung bzw. Generalisierung reserviert sind. Die Generalisierungsmenge wird zum Training überhaupt nicht berücksichtigt, um nachträglich eine vollständig unvoreingenommene Überprüfung der Güte zu ermöglichen. Die Güte auf der Validierungsmenge, falls vorhanden, wird im Laufe der Optimierung mehrfach zur Überprüfung verwendet: Die Güte auf der Validierungsmenge liefert ein unabhängiges Maß für den Fortschritt der Optimierung auf der Basis der Trainingsmenge und dient außerdem zur Vermeidung von Überanpassung.

Wahl der Faktoren

In der Ausführungsform besteht die Möglichkeit, Untermengen der Faktoren heranzuziehen, um beispielsweise Modelle für solche zukünftigen Muster zu gewinnen, für die nur diese Untermenge der Faktoren zur Verfügung steht.

Netzoptimierung

Simplex-Optimierung

Bei der Optimierung geht es um die Suche nach einem Maximum der Likelihoodfunktion, bezogen auf die Daten der Trainingsmenge. Der Parameterraum der Suche besteht aus den $n-K$ noch aktiven Netzgewichten zusammen mit den globalen Zeitkonstanten λ_{k0} , $k=1, \dots, K$. Daraus ergibt sich ein n -dimensionaler Raum, in dem die Suche erfolgt.

Die in der Ausführungsform implementierte Methode der Suche bedient sich der Konstruktion von einem n -fachen *Simplex* in diesem Raum nach der bekannten Methode von Nelder und Mead (1965). Die Suche erfordert die Bildung einer n -dimensionalen *Simplex* im Parameterraum. Eine Simplex läßt sich durch Vorgabe von $n+1$ nicht entarteten Ecken bestimmen, d.h., die entsprechenden Kanten sind alle linear unabhängig voneinander. Sie umfaßt somit eine n -dimensionale Punktwolke im Parameterraum. Die Optimierungssuche erfolgt in Epochen. Während jeder Epoche wird die Gütefunktion auf der Trainingsmenge an verschiedenen Stellen im Parameterraum ausgewertet, nämlich am aktuellen Ort sowie an n weiteren Ecken, die durch Zusammensetzung von Operationen wie Reflexion, Expansion/Kontraktion in einer Richtung, usw., definiert sind. Die Richtungen dieser Operationen werden auf der Basis der Werte der Gütefunktion an den zur vorigen Epoche definierten Ecken automatisch gewählt. Die Abnahme der Gütefunktion in der Ausführungsform ist monoton, und die Suche endet immer bei einem (wenigstens lokalen) Minimum.

Berücksichtigung der Validierungsmenge

Die oben beschriebene Validierungsmenge dient, falls vorhanden, zur Kontrolle des Optimierungsfortschritts und zur Vermeidung von Überanpassung.

Als Kennzahlen der augenblicklichen Güte der Optimierung bezüglich der Trainings- und Validierungsmengen werden in der Ausführungsform laufend die Größen *Minus Log-likelihood pro Muster* der beiden Mengen berechnet und ausgegeben. Während diese Kennzahl auf der Trainingsmenge monoton abnehmen muß, sind vorübergehende Schwankungen der entsprechenden Kennzahl auf der Validierungsmenge möglich, ohne dass eine Überanpassung bereits stattgefunden hat. Ein monotoner *Anstieg* der Kennzahl auf der Validierungsmenge sollte jedoch zum Abbruch der weiteren Optimierung und zu einer

Komplexitätsreduktion führen. Dieser Art der Abbruch stellt eine Art Notbremse für die Vermeidung von Überanpassung vor.

Ein mögliches Abbruchkriterium, das automatisch durchgeführt werden kann, wird durch die Führung der *exponentiell geglätteten Güte* der Validierungsmenge erreicht. Falls diese geglättete Kenngröße das bisherige Minimum des aktuellen Optimierungsschritts um einen festen Prozentsatz überschreitet (Verschlechterung der Güte), wird die Optimierung abgebrochen. Eine prozentuale Zunahme von etwa 1 % Toleranz wurde als Erfahrungswert für typische Größen der Trainingsmenge um die 300 oder mehr Datensätze festgestellt. Bei dieser Toleranz und bei etwa gleicher Größe von Trainings und Validierungsmengen wird das Training öfter durch die Erreichung eines Minimums auf der Trainingsmenge gestoppt als durch die Verschlechterung der Güte auf der Validierungsmenge. Dieser "normale" Abbruch ist bevorzugt, weil eine (fast) monotone Verbesserung der Güte auf der Validierungsmenge ein Zeichen darstellt, dass das neuronale Netz echte zugrunde liegenden Strukturen, und nicht einfach das Rauschen, erkannt hat.

Im Beispiel für die Ausführungsform wurde keine Validierungsmenge verwendet. Somit erfolgt der Abbruch allein aufgrund des Minimums auf der Trainingsmenge.

Strukturoptimierung und Komplexitätsreduktion

Aus der für die Ausführungsform beschriebenen Simplexoptimierung resultiert ein Satz von Gewichten $\{w_{[1]}, \dots, w_{[n]}\}$ und anderen Parametern, welche ein lokales Minimum der negativen Log-Likelihood bestimmen. (Die Numerierung $[1] \dots [n]$ der Gewichte in diesem Zusammenhang beinhaltet nicht die topologische Ordnung der Gewichte.) Dieses Minimum bezieht sich auf die feste Zahl n der Gewichte und eine feste Topologie. Um Überanpassung zu vermeiden, ist es wünschenswert, durch eine Ausdünnung der Gewichte die Komplexität zu verringern, soweit dies ohne einen signifikanten Verlust der Güte möglich ist.

Ausdünnung (Pruning) bezeichnet die Deaktivierung von Konnektoren. Dazu werden deren Gewichte auf einen festen Wert "eingefroren" (in der Ausführungsform null, wobei man auch vom "Entfernen" sprechen kann). Es ist im Prinzip möglich, einzelne Gewichte oder auch ganze Knoten zu entfernen. Im letzteren Fall werden alle Gewichte deaktiviert, die entweder in den zu entfernenden Knoten einführen oder aus dem Knoten weiterführen.

In der Ausführungsform wird eine Phase der Komplexitätsreduktion im Netz im Anschluß an einer Optimierungsphase (Simplex-Verfahren) durchgeführt. Der erste Schritt dazu ist die „Ausdünnung“ einzelner Konnektoren. Im Anschluß werden Kombinationen aus verschiedenen

Konnektoren im Hinblick auf Redundanz getestet. Schließlich wird die Konsistenz der Topologie überprüft und gegebenenfalls Konnektoren bzw. Knoten entfernt, die aufgrund der bisherigen Entfernung anderer Konnektoren und Knoten nicht mehr zur Aussage beitragen können. Diese Prozedur ist zwar nicht Gegenstand der Erfindung, gehört jedoch zur guten Praxis gemäß dem Stand der Technik.

Zur Komplexitätsreduktion werden in der Ausführungsform automatisch verschiedene statistische Hypothesen gebildet, die mittels eines Likelihood-ratio-tests bezüglich eines vorgegebenen Signifikanzniveaus überprüft werden. Bestimmte Gewichte bzw. Parameter werden hierbei als obligatorisch betrachtet werden, d.h., sie werden auf keinen Fall entfernt. Dazu gehören die globalen Zeitparameter λ_{ok} .

Rangfolge der Konnektoren

Um die Reihenfolge der zu überprüfenden Konnektoren festzustellen, wird in der Ausführungsform zunächst die Testgröße $\log(\text{Likelihood-Ratio})$ gebildet. Dabei stellt man sich für jedes Gewicht $w_{[A]}$ zwei Netzwerke vor:

- Das Netz mit allen aktuellen Gewichten (n Freiheitsgrade), einschließlich $w_{[A]}$.
- Das Netz mit allen aktuellen Gewichten außer $w_{[A]}$, das deaktiviert wird ($n-1$ Freiheitsgrade).

Beim Netz mit $w_{[A]}$ deaktiviert werden die übrigen Gewichte bei den aktuell optimierten Werten eingefroren.

Testing

Nachdem eine Rangfolge $\{w_{[1]}, \dots, w_{[n]}\}$ der Gewichte gemäß dem "Likelihood-ratio" bekannt ist, werden in der Ausführungsform die in dieser Reihenfolge zwecks der Ausdünnung getestet, bis maximal G_{\max} Gewichte zu entfernen sind. Angenommen, $A-1$ Gewichte sind bereits entfernt worden, kann man für das A -te zusätzliche Gewicht der Reihenfolge $w_{[A]}$ zwei Hypothesen testen.

- Teststatistik für die Hypothese H_{A-1} : Likelihood-Ratio für das Netz mit Gewichten $\{w_{[1]} \dots w_{[A-1]}\}$ deaktiviert ($n-A+1$ Freiheitsgrade)
- Teststatistik für die Hypothese H_A : Likelihood-Ratio für das Netz mit Gewichten $\{w_{[1]} \dots w_{[A]}\}$ deaktiviert ($n-A$ Freiheitsgrade)

Die Hypothese H_A wird nun zweimal getestet:

- H_A versus H_{A-1} und
- H_A versus H .

Die Signifikanz wird mit dem Chi-Quadrat-Test bezüglich des Likelihood-Ratios angewandt. Falls bei einer der beiden Vergleiche H_A angenommen wird (Ausdünnung von A ergibt eine signifikante Verschlechterung), wird der Konnektor A nicht entfernt, und der Ausdünnungsschritt wird beendet.

Beim Deaktivieren wird der Konnektor aus der Liste der aktiven Konnektoren entfernt und das zugehörige Gewicht eingefroren (meistens gleich null).

Während einer Ausdünnungsphase wird in der Ausführungsform die Zahl G der entfernten Konnektoren auf eine maximale Zahl $G_{\max}=n/10$ begrenzt, wobei n die Anzahl der verbleibenden Konnektoren ist.

Weitere Komplexitätsreduktion

In der Ausführungsform werden durch paarweise Analyse der Gewichte im Hinblick auf die Likelihood der Daten und unter Berücksichtigung verschiedener Korrelationseigenschaften weitere Konnektoren entfernt. Dieser Schritt ist jedoch nicht zwingend erforderlich für die Funktion des lernfähigen Modells und kann daher weggelassen werden. Ebenso ist es möglich, die Erfindung mit anderen Techniken zur Komplexitätsreduktion, die in verschiedenen lernfähigen Systemen eventuell bereits implementiert sind, zu verbinden.

Überprüfung der Topologie

Ausdünnung oder Entfernung einzelner Konnektoren kann zur Isolierung eines Knotens von Eingangssignalen, Ausgangssignalen, oder (im Falle eines versteckten Neurons) von beiden führen. In diesem Fall wird in der Ausführungsform ein Deaktivierungsflag für den Knoten gesetzt. Für Neuronen der Ausgangsschicht bedeutet z.B. "Isolierung", dass weder von der Eingangsschicht, noch von der versteckten Schicht aktive Konnektoren vorhanden sind. Falls alle Konnektoren von einem Eingangsneuron zur versteckten und zur Ausgangsschicht entfernt worden sind, ist der Bias der linearen Konnektoren auch zu deaktivieren.

Ein verstecktes Neuron, das von allen Eingängen isoliert worden ist, kann noch mit Ausgängen verbunden ist. Die "eingefrorenen" Beiträge solcher versteckter Neuronen zum Ausgang sind dann redundant, weil sie im Prinzip nur die Biaswerte der übrigen, aktiven Konnektoren

verändern. Daher werden solche Neuronen deaktiviert, und verbleibende Konnektoren zur Ausgangsschicht werden entfernt.

Die verschiedenen Überprüfungen können zur weiteren Isolierung von Knoten führen. Deshalb wird die Prozedur solange iteriert, bis die Topologie konstant bleibt.

Beenden des Trainings und Ausgabe

Wenn nach der letzten Simplex-Optimierung keine weitere Komplexitätsreduktion möglich ist, wird in der Ausführungsform das Training beendet. Alle Gewichte und sonstige Parameter erhalten ihre endgültigen Werte, die in hierfür angelegten Dateien gespeichert werden.

Somit ist das trainierte neuronale Netz eindeutig bestimmt. Durch Einlesen dieser gespeicherten Werte kann entweder gleich anschließend oder zu einem zukünftigen Zeitpunkt das trainierte neuronale Netz gemäß der obigen Beschreibung verwendet werden, um für beliebige Daten, welche die unabhängigen Faktoren („Kovariaten“) x enthalten, die Ausgabewerte und somit die oben definierten Funktionen $f_k(t)$, $\lambda_k(t)$, und $S_k(t)$, zu den Kovariaten x zu gewinnen. Mit diesen Funktionen ist das Wahrscheinlichkeitsmodell nun bestimmt.

Insbesondere ist es selbstverständlich möglich, den Verlauf dieser Funktionen in Abhängigkeit von gewählten Faktoren, auszurechnen. Eine solche abhängige Bestimmung ist zur Evaluierung der erwarteten Wirkung eines Therapiekonzeptes sinnvoll, wenn die zu bewertenden Therapien als „Faktoren“ zum Training verwendet worden sind.

Beispiel

Daten

Zur Veranschaulichung der Funktionsweise der Erfindung in der Ausführungsform wurden zunächst 1000 fiktive Patientendatensätze mit 9 Faktoren (Kovariaten) mittels eines Zufallsgenerators erzeugt. Die ersten 7 Faktoren wurden als Realisierungen einer multivariaten Gaußverteilung erzeugt. Hierfür wurden im Ausführungsbeispiel Mittelwerte und Varianzen der Faktoren und eine Kovarianzmatrix vorgegeben:

Faktor	xlypo	xer	xpr	xage	xtum	xupa	xpai
Mittelwert	0.50	0.45	0.45	0.50	0.51	0.50	0.50
Varianz	0.071	0.087	0.097	0.083	0.083	0.084	0.083

Die angenommene Kovarianzmatrix war

	xlypo	xer	xpr	xage	xtum	xupa	xpai
xlypo	1.00	-0.06	-0.09	0.03	0.42	0.02	0.05
xer	-0.06	1.00	0.54	0.29	-0.07	-0.18	-0.19
xpr	-0.09	0.54	1.00	0.03	-0.06	-0.07	-0.14
xage	0.03	0.29	0.03	1.00	0.04	0.02	0.00
xtum	0.42	-0.07	-0.06	0.04	1.00	0.03	0.06
xupa	0.02	-0.18	-0.07	0.02	0.03	1.00	0.54
xpai	0.05	-0.19	-0.14	0.00	0.06	0.54	1.00

Um eine möglichst realistische Situation darzustellen wurden diese Werte in der Größenordnung so gewählt, wie sie in der wissenschaftlichen Literatur im Falle des Mammakarzinoms für bestimmte Faktoren bekannt sind. Für die Funktion der Erfindung sind die genauen Annahmen sowie die Interpretation der Faktoren jedoch völlig unerheblich.

Neben den sieben benannten Faktoren wurden zwei weitere binäre Faktoren („Therapien“) „ct“ und „ht“ zufällig erzeugt. Für ht wurden zufällig 50 % der Wert 1, 50 % der Wert 0 zugeordnet. Im Ausführungsbeispiel für ct wurden nur 1 % der Wert 1, 99 % der Wert 0 zugeordnet. Somit ist zu erwarten, daß ct nicht im neuronalen Netz als Einflußfaktor erkannt wird.

Die ersten zehn der hieraus resultierenden Datensätze sind wie folgt:

Patienten- Nummer	xlypo	xer	xpr	xage	xtum	xupa	xpai	ct	ht
1	0.07	0.89	1.41	0.36	0.49	0.31	0.22	0	1
2	0.25	0.23	0.98	0.15	0.10	0.31	0.05	0	0
3	0.56	0.52	0.79	0.09	0.22	-0.22	-0.07	0	1
4	0.61	0.83	1.10	0.73	0.56	0.21	0.44	0	1
5	0.97	0.38	0.70	0.61	0.51	0.97	0.72	0	0
6	0.44	0.22	0.07	0.90	0.80	0.60	0.55	0	1
7	0.46	0.24	0.47	0.14	0.60	0.57	0.31	0	0
8	0.42	0.60	0.41	0.36	0.54	0.23	0.47	0	0
9	-0.01	0.22	0.80	0.52	0.38	-0.13	0.41	0	0
10	0.80	0.41	0.19	0.11	0.45	0.40	0.51	0	0

Für den Einfluß der Faktoren auf einen Krankheitsverlauf wurden zunächst drei unabhängige Risikowahrscheinlichkeiten $\text{risk}(i)$, $i=1,3$ erzeugt. Folgendes Modell wurde unterstellt:

$$\text{risk}(1)=\exp(r_1+r_2+r_3+r_4-r_h) \quad \text{risk}(2)=\exp(r_1+r_3+r_4) \quad \text{risk}(3)=\exp(r_1)$$

mit

$$\begin{aligned} r_1 &= 2 \text{ (xlypo-median(xlypo))} \\ r_2 &= 0,5 \text{ (xtum-median(xtum))} \\ r_3 &= 0,75 \text{ (xupa-median(xupa))} \\ r_4 &= 1,5 \text{ (xpai-median(paumed)) und} \\ r_h &= 1 \text{ falls ht=1.} \end{aligned}$$

Aus diesen Risikowerten wurden tatsächliche Versagenszeiten der drei Ausprägungen als zufällige Realisierungen einer Exponentialverteilung bzw. einer modifizierten Exponentialverteilung mit Zeitkonstante 200 Monate erzeugt. Für die 3. Ausprägung wurde zusätzlich angenommen, daß ein Versagen nach spätestens 24 Monaten möglich ist, um eine Situation mit konkurrierenden Risiken ähnlich dem Lokalrezidiv beim Mammakarzinom zu erzeugen. Diese Daten wurden gemäß einer simulierten „Studie“ zensiert, und eine „Beobachtung“ wurde nach dem Schema der unter Abbildung 1 dargestellten Priorität simuliert.

Aus dem im Ausführungsbeispiel unterstellten Modell geht hervor, daß für das Versagen der 3. Ausprägung nur der Faktor „xlypo“ kausal ausschlaggebend ist. Dennoch besteht ein *indirekter* Zusammenhang zwischen den übrigen Faktoren und den Beobachtungen der 3. Ausprägung, weil erhöhte Risiken der übrigen Faktoren die Wahrscheinlichkeit für eine Beobachtung des Versagens der 3. Ausprägung *senken* können. Diese Eigenschaft des unterstellten Modells ist zwar unerheblich für die Funktion der Erfindung, veranschaulicht aber einen typischen Nutzen.

Trainiertes neuronales Netz

Die Neuronen der Ausgangsschicht werden gemäß den Gleichungen 4 bis 7 und 10 mit $N_{\text{time}} = 1$ und $N_{\text{key}} = 3$ zugeordnet, so daß 3 Neuronen der Ausgangsschicht ursprünglich aktiv sind. Für die Ausführungsform werden jeweils 9 Neuronen der Eingangs- bzw. der versteckten Schicht ursprünglich aktiviert. Das nach der beschriebenen Methode trainierte neuronale Netz wird in der Figur 3 veranschaulicht („xpai“ und „xpai1“ sind identisch). Man beachte, dass zum Ausgang „O3“ nur ein Konnektor vorhanden ist, und zwar vom Knoten (Neuron) „xlypo“. Die Ausgänge O1 bis O3 sind den Risiken „risk(1)“ bis „risk(3)“ zugeordnet.

Eine vollständige eindeutige Darstellung des trainierten neuronalen Netzes erfolgt durch Angabe der verbleibenden Konnektoren mit ihren zugehörigen Gewichten und Bias-Werten sowie durch die Überlebenszeitskalen. Hierfür werden in der Tabelle 2a für jedes Neuron, zu

dem ein aktiver Konnektor führt („tgt“), alle Quellen („src“) mit den entsprechenden Gewichten („wt“) angegeben. Man beachte, daß viele Konnektoren nicht mehr aktiv sind.

tgt	src	wt	src	wt	src	wt	src	wt	src	wt	src	wt	src	wt	src	wt	src	wt
h1	ht	13.5																
h6	xlypo	0.53	xupa	-1.78	xtum	1.02												
h7	xer	1.98	xpr	-1.37														
h8	xage	1.70																
h9	xpr	2.31																
o1	h1	-1.70	h6	0.30	ht	-1.10	xlypo	0.19	xpai	0.72	xupa	0.63	xtum	0.22				
o2	h1	2.03	h6	-0.68	h7	-0.86	h8	0.33	h9	-0.64	xlypo	0.64	xpai1	0.91	xer	0.56	xage	-0.42
o3	xlypo	2.39																

Tabelle 2a

Die Bias-Werte sind wie in der Tabelle 2b angegeben:

ht	xlypo	xpai	xupa	xtum	ct	xer	xage	xpr	h1	h2	h3	h4	h5	h6	h7	h8	h9	o1	o2	o3
0.17	0.16	0	0	0	0	0	0	0	-0.94	0	0	0	0	0.86	1.31	0	2.07	1.03	0.66	-0.11

Tabelle 2b: Bias Werte (automatisch 0 für inaktive Neuronen)

Die Werte der für die Spezifikation des Modells der Gleichung 6 notwendigen Überlebenszeitskalen λ_{0k} sind schließlich der Tabelle 2c zu entnehmen (die Einheiten dieser Angaben entsprechen sinnvollerweise der obigen Zeitkonstante von 200 Monaten):

λ_{01}	λ_{02}	λ_{03}
0.53/200	0.13/200	0.27/200

Tabelle 2c

Zeitliche Variation

Um zeitlich variierende Ausgangsneuronen zu verwenden, könnte statt $N_{\text{time}} = 1$ wie hier verwendet ein höherer Wert eingesetzt werden. Die Anzahl der Ausgangsneuronen bestimmt sich dann aus der Gleichung 10.d. Im Falle $N_{\text{key}} = 3$ und $N_{\text{time}} = 2$ wäre dann z.B. $N_o = 6$. Das Training würde auf die bisher beschriebene Weise erfolgen. Die eventuellen zeitlichen Variationen der verschiedenen Ausprägungen könnten im Rahmen des Modells der Gleichungen 4 bis 7 unabhängig von einander bestimmt werden, die Aufgabe der Erfassung konkurrierender Risiken würde hierdurch insbesondere keinesfalls beeinträchtigt werden.

Patentansprüche

1. Verfahren zur Ermittlung konkurrierender Risiken für Objekte nach einem Anfangsereignis auf Grundlage bereits gemessener oder sonst objektivierbarer Trainingsdatensätze, in welchem mehrere aus einem lernfähigen System gewonnene Signale in einer objektiven Funktion so kombiniert werden, dass das lernfähige System die zugrunde liegenden Wahrscheinlichkeiten der jeweiligen konkurrierenden Risiken erkennen oder vorhersagen kann.
2. Verfahren nach Anspruch 1, in welchem für die Trainingsdatensätze gemessene oder in sonstiger Weise objektiv erfaßte Daten des Anfangsereignisses und einer Nachbeobachtung bis zu einem vorgegebenen Zeitpunkt verwendet werden.
3. Verfahren nach Anspruch 2, in welchem der letzte Zeitpunkt der Nachbeobachtung explizit in den Trainingsdatensätzen verwendet wird.
4. Verfahren nach einem der vorangegangenen Ansprüche, in welchem bei der Beobachtung einer Versagensausprägung zu einem Zeitpunkt die anderen Ausprägungen ausgeschlossen werden.
5. Verfahren nach einem der vorangegangenen Ansprüche, in welchem die objektive Funktion L in Abhängigkeit einer Funktion P gegeben ist:

$$L(\mu; \{x_j, t_j, \delta_{jk}\}) = \prod_{j=1}^n P\left(\left[f_{LS(k, x_j)}(t_j)\right] \left[S_{LS(k, x_j)}(t_j)\right] k=1, \dots, K\right),$$

wobei μ die Parameter des lernfähigen Systems, $f_{LS(k, x_j)}(t_j)$ die Versagensrate der Ausprägung k und $S_{LS(k, x_j)}(t_j)$ den Erwartungswert des Anteils an Objekten j mit beobachteten Merkmalen x_j , welche kein Versagen der Ausprägung k zum Zeitpunkt t_j erleiden, bezeichnen und P aufgrund des logischen Zusammenhangs aus δ_{jk} bestimmt wird, mit $\delta_{jk} = 1$, wenn ein Objekt j ein Versagen der Ausprägung k zum Zeitpunkt t_j erlitten hat und sonst $\delta_{jk} = 0$.

6. Verfahren nach Anspruch 5, in welchem

$$L(\mu; \{\mathbf{x}_j, t_j, \delta_{jk}\}) = \prod_{j=1}^n \prod_{k=1}^K [f_{LS(k, \mathbf{x}_j)}(t_j)]^{\varepsilon_{jk}} [S_{LS(k, \mathbf{x}_j)}(t_j)]^{\psi_{jk}}$$

als objektive Funktion verwendet wird, wobei ε_{jk} und ψ_{jk} aufgrund des logischen Zusammenhangs aus δ_{jk} bestimmt werden.

7. Verfahren nach Anspruch 6, in welchem

$$L(\mu; \{\mathbf{x}_j, t_j, \delta_{jk}\}) = \prod_{j=1}^n \prod_{k=1}^K [f_{LS(k, \mathbf{x}_j)}(t_j)]^{\delta_{jk}} [S_{LS(k, \mathbf{x}_j)}(t_j)]^{1-\delta_{jk}}$$

als objektive Funktion verwendet wird.

8. Verfahren nach einem der vorangegangenen Ansprüche, in welchem ein neuronales Netz als lernfähiges System verwendet wird.

9. Verfahren nach Anspruch 8, in welchem ein neuronales Netz der Architektur MLP (multi-layer Perzeptron) verwendet wird.

10. Verfahren nach einem der Ansprüche 1 – 7, in welchem das lernfähige System eine rekursive Partitionierung durchführt, wobei

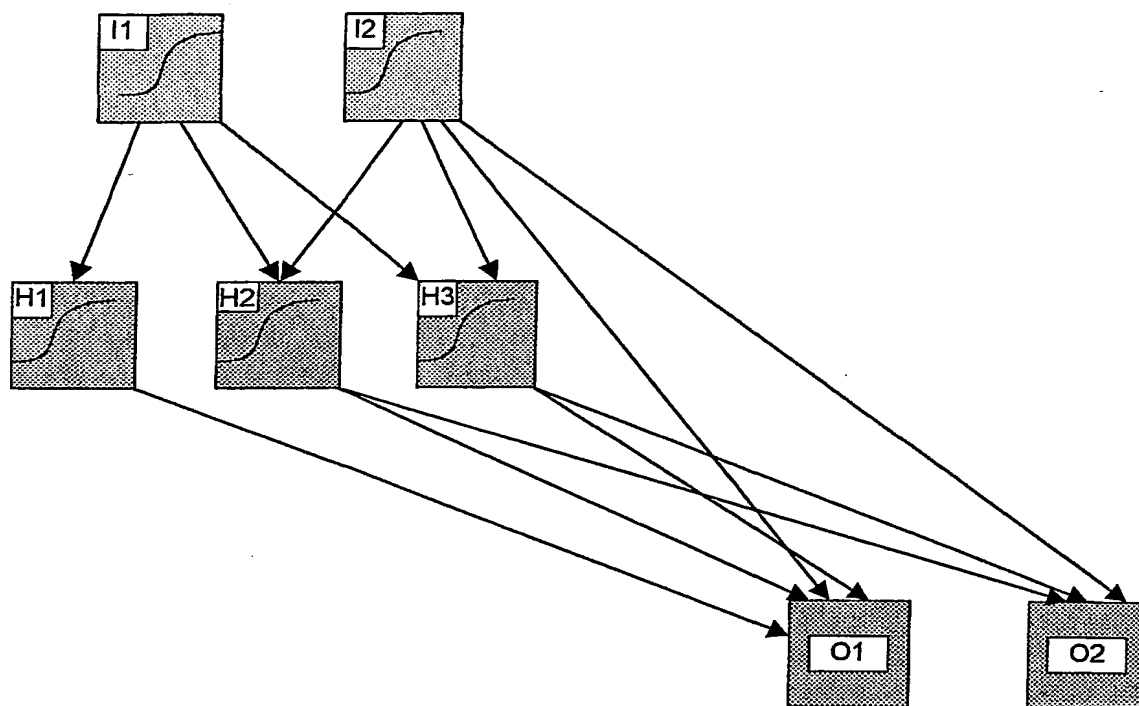
jedem Objekt einem Knoten zugeordnet wird,

jedem Knoten die Häufigkeiten oder Wahrscheinlichkeiten aller Ausprägungen zugeordnet wird, und

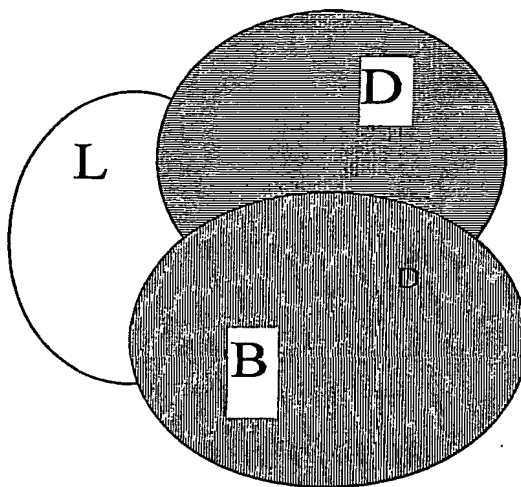
die Partitionierung so durchgeführt wird, dass die objektive Funktion optimiert wird, die diese Häufigkeiten oder Wahrscheinlichkeiten statistisch berücksichtigt.

11. Verfahren nach einem der vorangegangenen Ansprüche, in welchem das lernfähige System im Rahmen einer Entscheidungshilfe verwendet wird.

12. Verfahren nach einem der vorangegangenen Ansprüche, in welchem den verschiedenen Wahrscheinlichkeitsfunktionen der konkurrierenden Risiken Werte zur Ermittlung einer Strategie zugeordnet werden.



Figur 1



Figur 2

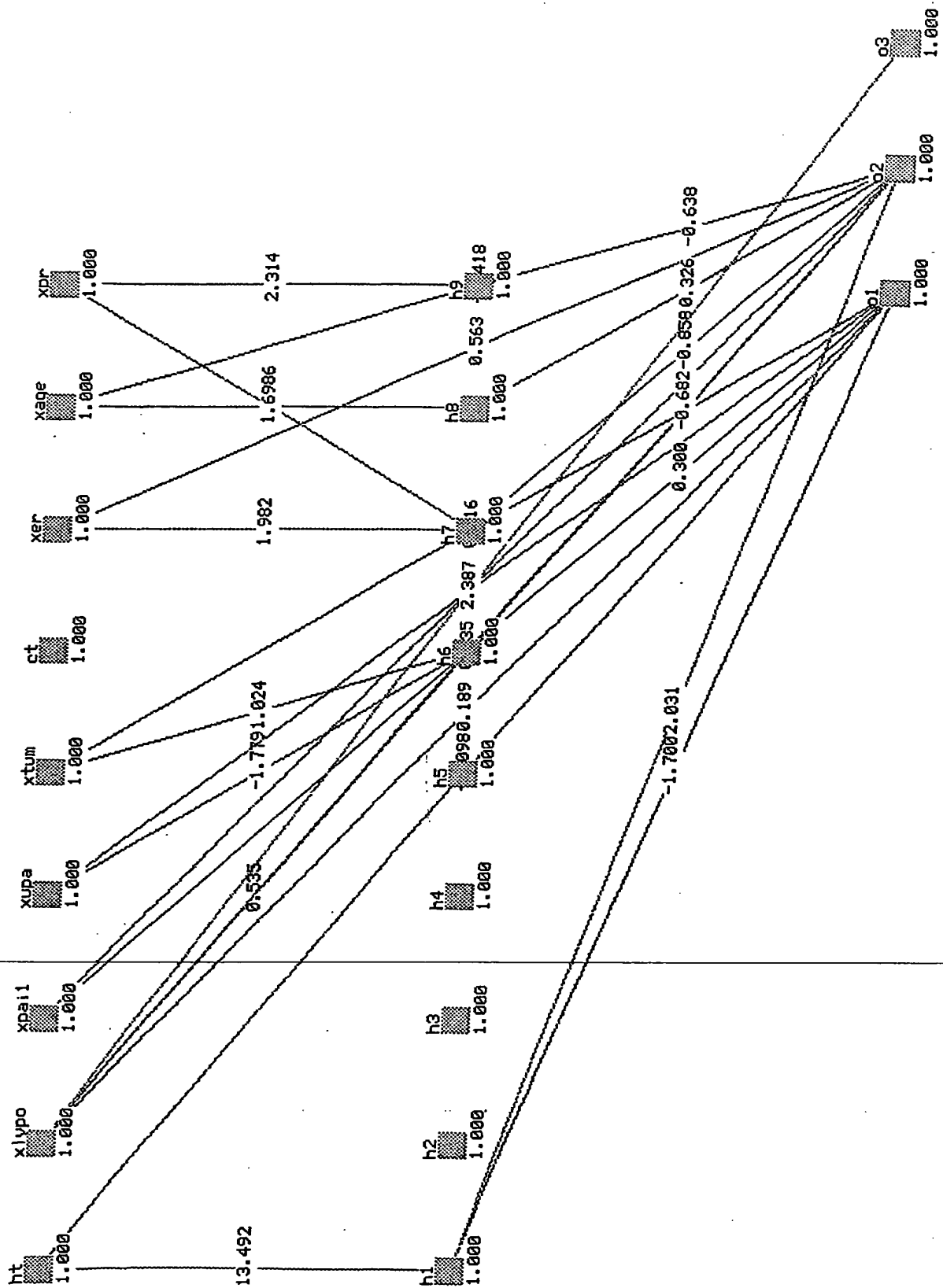


Figure 3

(12) NACH DEM VERTRAG ÜBER DIE INTERNATIONALE ZUSAMMENARBEIT AUF DEM GEBIET DES
PATENTWESENS (PCT) VERÖFFENTLICHTE INTERNATIONALE ANMELDUNG

(19) Weltorganisation für geistiges Eigentum
Internationales Büro



(43) Internationales Veröffentlichungsdatum
13. Juni 2002 (13.06.2002)

PCT

(10) Internationale Veröffentlichungsnummer
WO 02/047026 A3

(51) Internationale Patentklassifikation⁷: **G06N 3/02**

(21) Internationales Aktenzeichen: PCT/EP01/14411

(22) Internationales Anmeldedatum:
7. Dezember 2001 (07.12.2001)

(25) Einreichungssprache: Deutsch

(26) Veröffentlichungssprache: Deutsch

(30) Angaben zur Priorität:
100 60 928.7 7. Dezember 2000 (07.12.2000) DE

(71) Anmelder und

(72) Erfinder: KATES, Ronald, E. [US/DE]; Palnkamer Str.

49. 83624 Otterfing (DE). **HARBECK, Nadia** [DE/DE];
Palnkamer Strasse 49, 83624 Otterfing (DE).

(74) Anwalt: **WEIGELT, Udo**; Grünecker, Kinkeldey,
Stockmaier & Schwanhäusser, Anwaltssozietät, Maximil-
ianstrasse 58, 80538 München (DE).

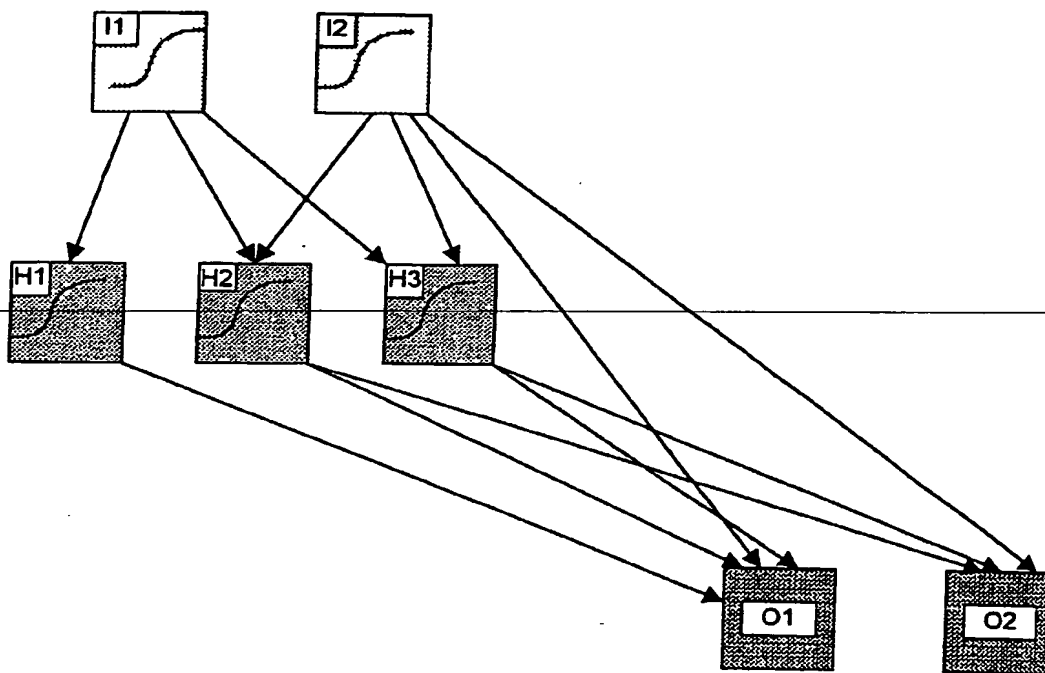
(81) Bestimmungsstaaten (*national*): AE, AG, AL, AM, AT,
AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR,
CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE,
GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR,
KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK,
MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU,
SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US,
UZ, VN, YU, ZA, ZM, ZW.

(84) Bestimmungsstaaten (*regional*): ARIPO-Patent (GH,
GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW),

[Fortsetzung auf der nächsten Seite]

(54) Title: METHOD FOR DETERMINING COMPETING RISKS

(54) Bezeichnung: VERFAHREN ZUR ERMITTLUNG KONKURRIERENDER RISIKEN



(57) Abstract: The invention relates to a method for determining competing risks for objects after an initial event based on already measured or otherwise objectifiable training data sets, in which a number of signals obtained from an adaptive system are combined in an objective function in such a manner that the adaptive system can identify or predict the underlying probabilities of the respective competing risks.

[Fortsetzung auf der nächsten Seite]

WO 02/047026 A3



eurasisches Patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), europäisches Patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI-Patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Veröffentlicht:

- mit internationalem Recherchenbericht
- vor Ablauf der für Änderungen der Ansprüche geltenden Frist; Veröffentlichung wird wiederholt, falls Änderungen eintreffen

(88) Veröffentlichungsdatum des internationalen**Recherchenberichts:**

6. November 2003

Zur Erklärung der Zweibuchstaben-Codes und der anderen Abkürzungen wird auf die Erklärungen ("Guidance Notes on Codes and Abbreviations") am Anfang jeder regulären Ausgabe der PCT-Gazette verwiesen.

(57) Zusammenfassung: Die Erfindung betrifft ein Verfahren zur Ermittlung konkurrierender Risiken für Objekte nach einem Anfangsereignis auf Grundlage bereits gemessener oder sonst objektivierbarer Trainingsdatensätze, in welchem mehrere aus einem lernfähigen System gewonnene Signale in einer objektiven Funktion so kombiniert werden, dass das lernfähige System die zugrunde liegenden Wahrscheinlichkeiten der jeweiligen konkurrierenden Risiken erkennen oder vorhersagen kann.

INTERNATIONAL SEARCH REPORT

Patent Application No
PCT/EP 01/14411A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 G06N3/02

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 7 G06N G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal, WPI Data, PAJ

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X A A	WO 91 18364 A (UNIV TEXAS) 28 November 1991 (1991-11-28) abstract page 7, line 28 -page 10, line 5 page 19, line 23 -page 23, line 24 page 34, line 40 -page 40, line 13 page 42, line 31 -page 48, line 39 claims 1,7,8 figures 1,4,5 ----- US 5 812 992 A (DE VRIES AALBERT) 22 September 1998 (1998-09-22) -----	1-3,8,9, 11,12 4-7,10

☐ Further documents are listed in the continuation of box C.☒ Patent family members are listed in annex.

* Special categories of cited documents:

- *A* document defining the general state of the art which is not considered to be of particular relevance
- *E* earlier document but published on or after the international filing date
- *L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- *O* document referring to an oral disclosure, use, exhibition or other means
- *P* document published prior to the international filing date but later than the priority date claimed

- *T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- *X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- *Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- *&* document member of the same patent family

Date of the actual completion of the international search

25 August 2003

Date of mailing of the international search report

29/08/2003

Name and mailing address of the ISA

European Patent Office, P.B. 5618 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax: (+31-70) 340-3016

Authorized officer

Jacobs, P

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/EP 01/14411

Patent document cited in search report		Publication date	Patent family member(s)	Publication date
WO 9118364	A	28-11-1991	US 5862304 A	19-01-1999
			AU 7980691 A	10-12-1991
			WO 9118364 A1	28-11-1991
<hr/>				
US 5812992	A	22-09-1998	NONE	
<hr/>				

INTERNATIONALER RECHERCHENBERICHT

Internationales Aktenzeichen

PCT/EP 01/14411

A. KLASSIFIZIERUNG DES ANMELDUNGSGEGENSTANDES
IPK 7 G06N3/02

Nach der internationalen Patentklassifikation (IPK) oder nach der nationalen Klassifikation und der IPK

B. RESEARCHIERTE GEBIETE

Recherchierte Mindestprüfstoff (Klassifikationssystem und Klassifikationssymbole)

IPK 7 G06N G06F

Recherchierte aber nicht zum Mindestprüfstoff gehörende Veröffentlichungen, soweit diese unter die recherchierten Gebiete fallen

Während der internationalen Recherche konsultierte elektronische Datenbank (Name der Datenbank und evtl. verwendete Suchbegriffe)

EPO-Internal, WPI Data, PAJ

C. ALS WESENTLICH ANGESEHENE UNTERLAGEN

Kategorie*	Bezeichnung der Veröffentlichung, soweit erforderlich unter Angabe der in Betracht kommenden Teile	Betr. Anspruch Nr.
X A A	<p>WO 91 18364 A (UNIV TEXAS) 28. November 1991 (1991-11-28) Zusammenfassung Seite 7, Zeile 28 -Seite 10, Zeile 5 Seite 19, Zeile 23 -Seite 23, Zeile 24 Seite 34, Zeile 40 -Seite 40, Zeile 13 Seite 42, Zeile 31 -Seite 48, Zeile 39 Ansprüche 1,7,8 Abbildungen 1,4,5</p> <p>--- US 5 812 992 A (DE VRIES AALBERT) 22. September 1998 (1998-09-22) -----</p>	<p>1-3,8,9, 11,12 4-7,10</p>

☐ Weitere Veröffentlichungen sind der Fortsetzung von Feld C zu entnehmen

☒ Siehe Anhang Patentfamilie

* Besondere Kategorien von angegebenen Veröffentlichungen :

A Veröffentlichung, die den allgemeinen Stand der Technik definiert, aber nicht als besonders bedeutsam anzusehen ist

E älteres Dokument, das jedoch erst am oder nach dem internationalen Anmeldedatum veröffentlicht worden ist

L Veröffentlichung, die geeignet ist, einen Prioritätsanspruch zweifelhaft erscheinen zu lassen, oder durch die das Veröffentlichungsdatum einer anderen im Recherchenbericht genannten Veröffentlichung belegt werden soll oder die aus einem anderen besonderen Grund angegeben ist (wie ausgeführt)

O Veröffentlichung, die sich auf eine mündliche Offenbarung, eine Benutzung, eine Ausstellung oder andere Maßnahmen bezieht

P Veröffentlichung, die vor dem internationalen Anmeldedatum, aber nach dem beanspruchten Prioritätsdatum veröffentlicht worden ist

T Spätere Veröffentlichung, die nach dem internationalen Anmeldedatum oder dem Prioritätsdatum veröffentlicht worden ist und mit der Anmeldung nicht kollidiert, sondern nur zum Verständnis des der Erfindung zugrundeliegenden Prinzips oder der ihr zugrundeliegenden Theorie angegeben ist

X Veröffentlichung von besonderer Bedeutung, die beanspruchte Erfindung kann allein aufgrund dieser Veröffentlichung nicht als neu oder auf erfinderischer Tätigkeit beruhend betrachtet werden

Y Veröffentlichung von besonderer Bedeutung, die beanspruchte Erfindung kann nicht als auf erfinderischer Tätigkeit beruhend betrachtet werden, wenn die Veröffentlichung mit einer oder mehreren anderen Veröffentlichungen dieser Kategorie in Verbindung gebracht wird und diese Verbindung für einen Fachmann naheliegend ist

Z Veröffentlichung, die Mitglied derselben Patentfamilie ist

Datum des Abschlusses der internationalen Recherche

25. August 2003

Absendedatum des internationalen Recherchenberichts

29/08/2003

Name und Postanschrift der internationalen Recherchenbehörde
Europäisches Patentamt, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax: (+31-70) 340-3016

Bevollmächtigter Bediensteter

Jacobs, P

INTERNATIONALER RECHERCHENBERICHT

Angaben zu Veröffentlichungen, die zur selben Patentfamilie gehören

Internationales Aktenzeichen

PCT/EP 01/14411

Im Recherchenbericht angeführtes Patentdokument	Datum der Veröffentlichung	Mitglied(er) der Patentfamilie	Datum der Veröffentlichung
WO 9118364 A	28-11-1991	US 5862304 A AU 7980691 A WO 9118364 A1	19-01-1999 10-12-1991 28-11-1991
US 5812992 A	22-09-1998	KEINE	

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

☒ **BLACK BORDERS**

☐ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**

☐ **FADED TEXT OR DRAWING**

☐ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**

☐ **SKEWED/SLANTED IMAGES**

☐ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**

☐ **GRAY SCALE DOCUMENTS**

☐ **LINES OR MARKS ON ORIGINAL DOCUMENT**

☒ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**

☐ **OTHER:** _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.